

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:17 ; Search time 170.72 Seconds
(without alignments)
17.102 Million cell updates/sec

Title: US-09-331-631A-1_COPY_74_116
Perfect score: 248
Sequence: 1 NQEDPQTECQCQCRRCROE.....RQOYCRCKEICEEEY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_66:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	47.2	566	2 S22477	vicilin precursor
2	105	42.3	588	1 FWCNAB	alpha-globulin B p
3	104	41.9	509	2 S08059	alpha-globulin typ
4	96	38.7	411	2 T29475	hypothetical prote
5	94	37.9	605	2 S06398	alpha-globulin typ
6	80.5	32.5	425	2 T18592	hypothetical prote
7	80.5	32.5	600	2 T18593	hypothetical prote
8	80.5	32.5	810	2 T44430	protein PVI00 limp
9	79	31.9	572	2 T29880	hypothetical prote
10	78	31.5	47	2 JCS557	hypothetical prote
11	77.5	31.2	242	2 T29699	arginine/glutamate
12	74	29.8	388	2 T31887	hypothetical prote
13	74	29.8	388	2 T31888	hypothetical prote
14	74	29.8	438	2 T31889	hypothetical prote
15	74	29.8	445	2 T31898	hypothetical prote
16	73.5	29.6	335	2 T31560	hypothetical prote
17	73.5	29.6	335	2 T31561	hypothetical prote
18	72.5	29.2	330	2 T25169	hypothetical prote
19	71.5	28.8	335	2 T31559	hypothetical prote
20	70	28.2	1513	2 T23681	hypothetical prote
21	68.5	27.6	654	2 T30136	hypothetical prote
22	68	27.4	419	2 T18450	hypothetical prote
23	67	27.0	154	2 T27967	hypothetical prote
24	66.5	26.8	273	2 T20930	hypothetical prote
25	66.5	26.8	314	2 T27686	hypothetical prote
26	66.5	26.8	356	2 S31574	hypothetical prote
27	66.5	26.8	637	2 S35221	globulin begl prec
28	65	26.2	111	2 T20452	hypothetical prote
29	65	26.2	910	2 A34721	androgen receptor

30	65	26.2	911	2 B34721	androgen receptor
31	65	26.2	919	2 A39248	androgen receptor
32	65	26.2	1306	2 T13592	hypothetical prote
33	63.5	25.6	33	2 A41822	antimicrobial pept
34	63.5	25.6	51	1 HSMSS1	protamine - mouse
35	63.5	25.6	51	2 S03997	protamine 1 - rat
36	62.5	25.2	342	2 J07110	brain-specific mem
37	62	25.0	648	1 J01150	protein kinase (EC
38	61	24.6	1390	2 T14004	trif protein - sil
39	60	24.2	223	1 KASHS2	alpha-s2-casein pr
40	60	24.2	1969	2 T08875	histidine kinase h
41	60	24.2	4957	2 T03455	ALR protein - huma
42	60	24.2	5262	2 T03454	ALR protein - huma
43	59.5	24.0	294	2 T23682	hypothetical prote
44	59.5	24.0	1282	2 JF0120	glycoprotein A - m
45	59	23.8	285	1 I46207	involucrin - dog

ALIGNMENTS

RESULT 1
S22477
vicilin precursor - cacao
C:Species: Theobroma cacao (cacao)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L.; Fritz, P.J.
Plant Mol. Biol. 18, 1173-1176, 1992
A>Title: Comparison of the structure and nucleotide sequence of vicilin genes of coco
A:Reference number: S22477; MGIID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCH>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1.24/Domain: signal sequence #status predicted <Sig>
F:25-566/Product: vicilin #status predicted <MNT>

Query Match 47.2%; Score 117; DB 2; Length 566;
Best local similarity 50.0%; Pred. No. 4.2e-05;
Matches: 20; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 2 QEDPQTECQCQCRRCROESGPRQOYCRCKEICEEE 41
DB 78 EEELQRYQOCQCRCEQDQGOGRQOOCQCRCKEYKEE 117

RESULT 2

FWCNAB
alpha-globulin B precursor (clone C72) - upland cotton
N:Alternate names: seed storage protein; vicilin precursor
C:Species: Gossypium hirsutum (upland cotton)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A>Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA
A:Residues: 1-588 <CHL>
A:Cross-references: GB:M16891; NID:q167374; PIDN:AAA33071.1; PID:q167375
A:Experimental source: var. Coker 201
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed
A:Reference number: S06398
A:Accession: S06911
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-75/Domain: signal sequence #status predicted <IG>
F:26-588/Product: alpha-globulin storage proprotein #status predicted <MAI>
F:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	42.3%;	Score 105;	DB 1;	Length 588;
Best Local Similarity	47.5%;	Pred. NO. 0.00074;		
Matches 19;	Conservative 9;	Mismatches 10;	Indels 2;	Gaps 1;

```

OY      3 EPDTECCGCCRCRROESGPRQOQYCRCKCEICEEEE 42
      ||| ::||: |||| ||: |||| : ||:|:
DB      81 EDPQRREYECQGECRQGE--NQRPCCQGRCLKRFEEQ 118

```

RESULT 3
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant MOL. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed
A:Reference number: S06398
A:Accession: S08059
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-509 <CHL>
C:Superfamily: glycinin

Query Match	41.9%;	Score 104;	DB 2;	Length 509;
Best Local Similarity	48.7%;	Pred. No. 0.00084;		
Matches 19: Conservative	8;	Mismatches 10;	Indels 2;	Gaps 1;

```

QY      4 DPQTECCQCRRCRQESGPRQOQYCCRRCKEICEEEEE 42
      III ::::: IIIII III IIIII :` III::
Db      1 DPQRRYECQCECRQEE--RQDPQCCQCRCKRFEQEQ 37

```

RESULT 4
T29475
hypothetical protein T01D1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29475
R:Bradshaw, H.: Wohldmann, P.
Submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid T01D1.
#Reference number: T20623

A:Gene: CESP:T01D1.6
A:Map position: 2
A:Introns: 25/3; 304/3
A:Superfamily: glnadin
A:Experimental source: strain Bristol NZ; clone T01D1
C:Genetics:
A:Cross-references: EMBL:U080455; PIDD:AA837887.1; GSPDB:GN00020; CESP:T01D1.6

Query Match 38.7%; Score 96; DB 2; Length 411;
Best Local Similarity 39.5%; Pred. No. 0.0048;
Matches 17; Conservative 11; Mismatches 11; Indels 4; Gaps 2;

QY 2 QEDPQTECQ-CQRCRQESGPRQQYQQRCKETICEEERY 43
138 QCPAQDCCQDCCDCAACPPQQ---QPPQQCCQDCCCTTCQSDQY 17

RESULT

alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental Biochemistry of cottonseed embryogenesis and germination. XIX
A:Reference number: S06398
A:Accession: S06398
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-605 <CHL>
C:Superfamily: glycinin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match	37.9%	Score 94;	DB 2;	Length 605;
Best Local Similarity	35.9%	Pred. No. 0.01;		
Matches 14;	Conservative 12;	Mismatches 13;	Indels 0;	Gaps 0;

QY 2 QEDPTECQQCRRCRQGESBPRQQCYCRKCEICEE 40
| : | : : : | : | : | : : | : | :
Db I15 QQDPDKOFKECCORCWOMEQRPERKKOQCVCREDOED 15

RESULT 6

hypothetical protein AC3.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18592
R:McMurry, A.
Submitted to the EMBL Data Library, April 1996
A:Reference number: Z18995
A:Accession: T18592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <wild>
A:Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
A:Experimental source: clone AC3
C:Genetics:
A:Gene: CESP:AC3.3
A:Map position: 5
A:Introns: 18/3
A:Superfamily: q1iadin

Query Match	32.5%	Score 80.5;	DB 2;	Length 425;
Best Local Similarity	27.9%	Pred. No. 0.19;		
Matches 17; Conservative	9;	Mismatches 12;	Indels 23;	Gaps 2;

```
QY 2 QEDPQTECQOCQRRC-----QGESP-----RQOQYCARCKEICE 38
      1: :||||| 1: :||| 1: :||| 1: :||| 1:
Db 255 QDSAPQCCQCCQNTCCQQAAPYCCQCCQCAFPCCQCCQGSAPACQCCQCTSCQFTCCQCCQCTPCCQ 31
```

QY	39	E	39
Db	315	Q	315

